IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 November 2023, accepted 8 December 2023, date of publication 12 December 2023,
date of current version 18 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3341889

RESEARCH ARTICLE

Classification of Photovoltaic Faults Using PSO-Optimized Compact Convolutional Transformer

YING-YI HONG[1], (Senior Member, IEEE), LI-FAN CHEN[1], AND WEINA ZHANG[2]
[1]Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan City 32023, Taiwan
[2]College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 201306, China

Corresponding author: Ying-Yi Hong (yyhong@ee.cycu.edu.tw)

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant NSTC 112-2218-E-008-011.

ABSTRACT Diagnosing photovoltaic (PV) farms has become increasingly complex due to their large-scale presence in diverse environmental conditions. A comprehensive diagnosis process comprises four essential steps: detection, localization, classification, and remedy. This study primarily focuses on the classification of faults in grid-connected PV arrays, including line-to-line faults, open circuit faults, short circuit faults, and partial shading conditions. The deep learning-based Compact Convolutional Transformer (CCT) is employed for classifying these PV faults. To eliminate the need for heuristic parameter/hyperparameter tuning of the CCT model, this paper utilizes Particle Swarm Optimization to optimize parameters/hyperparameters such as kernel size, pooling size, stride, padding, number of multi-heads, and the number of transformer encoders. Given that CCT operates based on images, the study investigates the use of heat maps incorporating different sizes, DC/AC signals, and the number of fault signal cycles as inputs. To reduce the training dataset for CCT, Taguchi experiments are employed to generate orthogonal data while considering variations in irradiance and temperature. A realistic PV array subset is used to demonstrate the performance of the proposed method. The simulation results reveal that the proposed approach outperforms classical machine learning algorithms (Support Vector Machine, Decision Tree, K-Nearest Neighbor, and Random Forest) as well as convolutional neural network (CNN)-based models (AlexNet, Googlenet, ResNet50, VGG16, and VGG19). Specifically, the proposed method achieves the highest testing accuracy (97.34%) and ResNet50 exhibits the second best testing accuracy (93.237%) among all CNN-based the methods while Random Forest demonstrates the highest testing accuracy (84.24%) among classical machine learning methods.

INDEX TERMS Classification, compact convolutional transformer, diagnosis, photovoltaics, particle swarm optimization, Taguchi experiment.

I. INTRODUCTION

A. BACKGROUND OF PROBLEM

In 2022, the surface temperature was 0.86 °C higher than the 20th-century average of 13.9 °C and 1.06 °C higher than the period between 1880 and 1900 [1]. The primary source of the increased carbon dioxide (CO_2) emissions responsible for this rise in surface temperature is the utilization of fossil fuels, including oil, coal, and natural gas. To address the greenhouse effect and combat climate change, renewable energies are becoming increasingly crucial as alternative sources of power generation. Among these renewable options, photovoltaic (PV) power generation stands out as a clean and sustainable energy resource. However, the scale of PV farms has significantly expanded to support power generation in large-scale power systems. This expansion has led to challenges in the operation and control of the power grid, as well as increased complexity in the maintenance of PV installations.

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo.

B. REVIEW OF EXISTING WORKS

The classification of PV (photovoltaic) faults can be categorized into two main approaches: Virtual and Thermal

Methods (VTM) and Electrical-based Methods (EBM). VTMs are particularly well-suited for identifying issues such as module breakage and browning, which primarily manifest as physical and thermal changes. On the other hand, EBMs are instrumental in diagnosing electrical faults occurring within PV systems. EBMs encompass various techniques, including electric signal analysis, classical intelligent methods, deep learning algorithms (DLA), and hybrid methods. Given that this work focuses on the classification of PV faults, the subsequent paragraphs will provide a detailed review of research related to EBMs.

(a) Electric signal analysis: Aboshady and Taha utilized measured differential currents across PV strings to classify short-circuit fault types [2]. Karmakar and Pradhan utilized the Thevenin equivalent resistance of the PV system to classify the type of PV fault [3]. Li et al. applied a voltage variation threshold to detect and classify partial shading conditions [4]. These approaches demonstrate how electric signal analysis can be a valuable tool for identifying and categorizing various PV faults based on electrical characteristics and measurements.

(b) Classical intelligent methods: They often based on machine learning and fuzzy theory, have been applied in several studies for the classification of PV faults: Xu et al. utilized the fuzzy c-means algorithm to classify fault types by considering parameters such as short-circuit current, maximum power point current, maximum power point voltage, and open-circuit voltage of a PV module [5]. Ul-Haq et al. employed a multi-layer feedforward neural network to classify short circuits and module mismatches using electrical signals and weather data [6]. Zhu et al. used three radial basis function-based neural networks to classify PV fault types by analyzing features of PV modules [7]. Laurino et al. employed a 4-layered neural network to classify PV faults based on current and voltage signals [8]. Li et al. discussed the effectiveness of five machine learning methods, including neural networks, support vector machine (SVM), decision trees (DT), random forests (RF), k-nearest neighbors (kNN), and naive Bayesian classifiers, for PV fault classification [9]. Badr et al. used Bayesian Optimization to fine-tune the hyperparameters of three fault classifiers (DT, kNN, and SVM) [10]. Eskandari et al. used three classifiers (SVM, naive Bayes, and logistic regression (LR)) to identify line-line and line-ground faults using features derived from the IV (current-voltage) curve of a PV system [11]. Rao et al. utilized nine features derived from the IV curves of a PV module as inputs to a pruned neural network, which incorporated a concrete dropout ratio to classify PV faults [12]. Chen et al. employed a semi-supervised ladder network to extract features from real-time faulty voltage and current data and classify various types of PV faults [13]. Dhibi et al. used an ensemble learning method to combine the classification results obtained by DT, kNN, and SVM classifiers [14]. These classical intelligent methods demonstrate the diverse range of techniques and algorithms applied to PV fault classification, highlighting the potential of machine learning and artificial intelligence in improving fault detection and classification in PV systems.

(c) Deep learning networks: Alves et al. used CNNs to classify anomalies in PV arrays, utilizing thermographic images as the input data [15]. Aziz incorporated weather data, seven features related to PV modules, and three converter data to generate two-dimensional scalograms. These scalograms were then employed as inputs to a fine-tuned AlexNet CNN for the purpose of classification [16]. Gao and Wei utilized voltage, current, and weather data as inputs to a two-dimensional CNN cascaded with a residual-gated recurrent unit to classify PV faults [17]. These studies demonstrate the application of deep learning and CNNs for fault classification in PV systems, with an emphasis on leveraging different types of data, including thermographic images, weather data, and electrical measurements.

(d) Hybrid methods: Hybrid methods involve the consolidation of multiple techniques to address complex problems. Several studies have utilized hybrid approaches for the classification of PV faults: Alrifaey et al. employed a stacked autoencoder to extract features from PV fault data obtained through wavelet transform. They then utilized long short-term memory (LSTM) for the classification of PV faults [18]. Eskandari et al. incorporated the Lasso penalty to select relevant features from fault signals. They used a weighted voting approach to ensemble the classification results obtained by LR, kNN, and SVM, with these models optimized using a genetic algorithm [19]. Guo et al. compared the decline in power generation caused by PV faults with normal operating conditions. They accomplished this through modeling using k-means clustering and transfer learning with LSTM, ultimately leading to the classification of PV faults [20]. These hybrid methods showcase the integration of various techniques, such as feature extraction, optimization, and machine learning, to enhance the accuracy and effectiveness of PV fault classification.

C. DEMERITS OF EXISTING WORKS

While the aforementioned studies are capable of classifying various types of PV faults, they also exhibit one or more limitations, which are described as follows:

(a) One limitation is the incomplete modeling of the bypass or blocking diode. In practice, these diodes are often implemented to prevent reverse current or reduce the mismatch fault currents [2], [3], [9], [10], [11], [13], [16].

(b) Another limitation is the reliance on heuristics and threshold values, such as a current threshold [2], Thevenin Equivalent Resistance [3], power loss threshold [4], voltage variation threshold [4], fuzziness degree [5], or decomposition level [18], to detect and classify faults or partial shading conditions. These heuristics may not account for the dynamic nature of power systems, which can undergo changes in short-circuit capacity at the point of common coupling.

(c) The studies often focus on a limited number of specific conditions, such as partial shading [4], line-line and line-ground faults [11], or open circuit and line-line

faults [19]. However, there are numerous mismatch conditions that may lead to difficulties in distinguishing PV faults. This is because some of these faults can produce similar voltage-current (VI) curves, making it challenging to differentiate between them.

(d) The methods employed in these studies are limited in their learning capability and may be prone to failure because PV systems are dynamic and constantly changing [1], [2], [3], [4], [5].

(e) The structure parameters or hyperparameters of classical neural networks [6], [7], [8], [12], [13], machine learning models [9], [11], [14], or deep learning networks [15], [16], [17], [18], [20] were not thoroughly explored and were determined through trial-and-error, which can be suboptimal and time-consuming.

D. MOTIVATIONS AND CONTRIBUTIONS OF PROPOSED METHOD

The advent of transformer models, a deep learning architecture built upon the parallel multi-head attention mechanism, has revolutionized various applications in advanced computer vision and natural language processing [21]. However, it's essential to note that transformer models often come with extremely large parameter sizes and substantial requirements for training data. This can be particularly challenging in domains where research data is limited, as is the case with realistic PV fault problems. Recently, a solution known as the Compact Convolutional Transformer (CCT) has demonstrated its ability to mitigate overfitting and surpass state-of-the-art convolutional neural networks (CNNs) when working with small datasets in computer vision applications. CCT achieves this through the right size and convolutional tokenization, making it a valuable advancement in addressing data limitations [22].

Motivated by the limitations identified in existing studies and the advancements in deep learning algorithms, this paper introduces a novel method that utilizes the CCT for the classification of PV fault types.

The main contributions of this paper can be summarized as follows:

(a) This paper addresses a broader range of PV fault types, including shorted circuits, open circuits, line-line faults, and partial shading conditions, in addition to the normal operating condition (in total 10 classes).

(b) The proposed method extracts features from faulty signals using convolutional layers and multi-head attentions within the CCT. Consequently, there is no need for additional feature extraction approaches or threshold setting, simplifying the classification process.

The novelties of this paper are summarized as follows:

(a) The proposed CCT model is trained using orthogonal datasets, which are minimized to meet the required dataset size. Consequently, this approach reduces computational time and the storage needed for training, making it more efficient and resource-friendly.

(b) Parameters and hyperparameters of the CCT are optimized using particle swarm optimization (PSO). This approach is employed to circumvent the need for heuristic methods and threshold setting, enhancing the efficiency and effectiveness of the CCT model.

(c) The study investigates heatmaps generated by both AC and DC faulty signals, while also considering the number of cycles of these faulty signals as important factors for the classification of PV faults.

This paper is organized as follows. Section II will provide an overview of the problem under investigation, detailing various types of faults in PV systems. Section III will delve into the background of the CCT. Section IV will introduce the proposed method, encompassing Taguchi's orthogonal experiments, the CCT model, and the implementation of PSO. Section V will present the simulation results, demonstrating the model's performance. Section VI will offer conclusions and insights into potential directions for future research.

II. STUDIED PV SYSTEM AND TYPES OF PV FAULTS

The studied PV system has a capacity of 81.6 kW, consisting of 24 PV strings connected in parallel. Each PV string is capable of generating a maximum power output of 3.4 kW and is equipped with its dedicated DC/DC converter, DC/AC inverter, and blocking diode. Additionally, each PV module is equipped with its own bypass diode. Given that all the PV strings are identical, the study focuses on two adjacent PV strings, as depicted in Fig. 1. Each PV string is composed of 10 modules. The AC voltage of the inverter is set at 3-phase 380 V, which is the secondary side of a 3-phase 380 V/11.4 kV transformer. The characteristics of each module are detailed in Table 1, providing key specifications for the PV modules used in the system.

TABLE 1. Characteristics of PV module.

Parameters	Quantities
Maximum Power (W)	340.0
Voltage at Maximum Power (V)	33.70
Current at Maximum Power (A)	10.09
Open circuit Voltage (V)	40.30
Short circuit Current (A)	10.56

This paper investigates various types of PV faults, including:

(a) Open modules in one or two strings (OMS): This occurs when some modules have open circuits, but bypass diodes still provide a circuit.

(b) Shorted modules in one or two strings (SMS): This involves some modules being short-circuited, such as when a bypass diode is shorted.

(c) One or two open strings (OS): This denotes an outage in the electrical pathway, resulting in zero measured power in the affected string.

(d) One or two shorted strings (SS): In this case, the entire string is grounded.

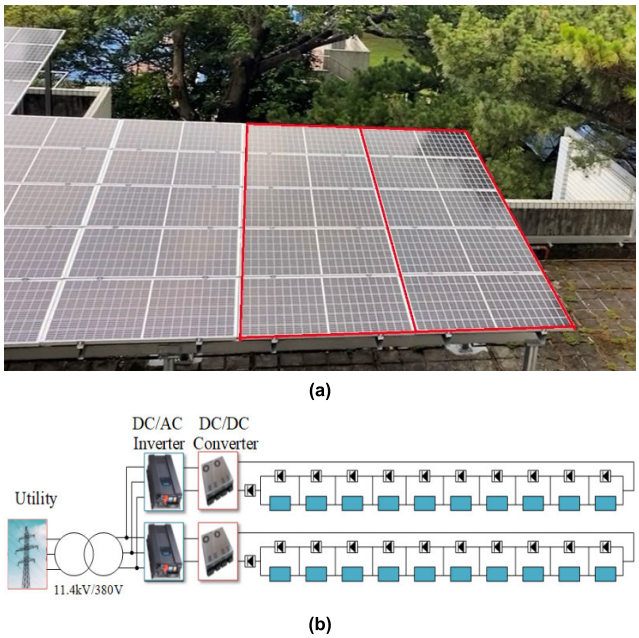


FIGURE 1. Studied two PV strings: (a) two PV strings marked in red color, (b) schematic diagram with 2 inverters and 2 converters for these 2 PV strings.

(e) Line-to-line fault (LL): Any terminal of a module in a PV string makes contact with any terminal of a module in an adjacent PV string.

(f) Partial shading (PS): This can be static, caused by factors like leaves, or dynamic, caused by moving clouds.

Actually, there are other types of PV faults, such as parameter deviation faults, degradation, diode faults, or inverter faults. Here are reasons why the studied OMS, SMS, OS, SS, LL, and PS conditions are concerned and may be prioritized over others:

Relevance to common issues: OMS, SMS, OS, SS, LL, and PS conditions are among the most common and critical issues in PV systems. Addressing these issues is crucial for the widespread adoption of solar energy [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20].

Practical significance: Certain faults, like parameter deviation faults, degradation, diode faults, or inverter faults, may occur less frequently in real-world scenarios. While these faults are undoubtedly important and can impact system performance, researchers may choose to address the more common and immediate concerns first before delving into more intricate issues.

It's essential to note that the focus on specific faults doesn't diminish the importance of considering a broader range of faults in the long run. As the field advances, various faults to enhance the overall reliability and performance of photovoltaic systems will be addressed and explored.

III. BACKGROUND OF COMPACT CONVOLUTIONAL TRANSFORMER (CCT)

The transformer model is known to have limitations related to the "data hunger" problem, requiring large amounts of training data. In contrast, convolutional neural networks (CNNs) are often preferred for smaller datasets due to their computational efficiency and memory requirements when compared to transformer models [10]. The CCT aims to bridge the gap between CNN and transformer model architectures. CCT is capable of attending to critical features within images while maintaining spatial invariance, particularly in scenarios with fewer interactions and weight sharing. Figure 2 illustrates the architecture of the CCT model [22].

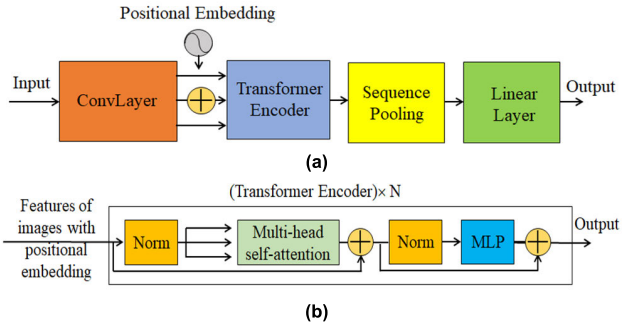


FIGURE 2. Compact convolutional transformer: (a) overall architecture, (b) transformer encoder.

Essentially, CCT is composed of convolutional layer (ConvLayer), positional embedding, transformer encoder, sequence pooling and linear layer (see Fig. 2(a)). Detailed description of CCT is given as follows.

(a) ConvLayer: The input is a sequence of vectors, called tokens, which will be converted into images. The ConvLayer block consists of a single convolution, ReLU activation, and a max pool. Given an image or feature map $\boldsymbol{x} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$, the ConvLayer operation is defined as follow:

$$x_0 = \mathrm{MaxPool}(\mathrm{ReLU}(\mathrm{Conv2d}(x))) \tag{1}$$

where the Conv2d operation has d filters [22]. The convolution and max pooling operations can be repeated.

(b) Positional embedding: Positional embedding augments spatial information into the tokens. This extra information is either a learned or sinusoidal embedding.

(c) Transformer encoder: Each transformer encoder comprises a multi-head self-attention (MHSA) layer, a multi-layer perceptron (MLP) block and layer normalization (denoted as "Norm" in Fig. 2(b)), GELU activation, and dropout. The transformer encoder can be repeated N times.

(d) Sequence pool: It is an attention-based method which pools over the output sequence of tokens. This generates crucial weights for the input tokens, which is applied as follows:

$$z = \boldsymbol{x}_N^t \boldsymbol{x}_N = \mathrm{Softmax}(g\,(f\,(\boldsymbol{x_0}))^t) \times f\,(\boldsymbol{x_0}) \tag{2}$$

where $f(\cdot)$ and $g(\cdot)$ are an N-layer transformer encoder and a linear layer operation, respectively. This output z can be sent via a classifier. One main difference between traditional vision transformer and new CCT is that the ''class'' token is substituted by this sequence pool.

IV. PROPOSED METHOD

The novelties of the proposed method include: (a) the establishment of training datasets using Taguchi's orthogonal experiments, (b) the optimization of hyperparameters and parameters of CCT using particle swarm optimization (PSO), (c) the conversion of various faulty signals using heatmaps. Additionally, the proposed CCT model will present the border PV fault classes and feature extractions.

A. TRAINING DATASETS GENERATED BY TAGUCHI'S METHOD

As training datasets for this work are typically scarce in practice, the approach taken involves using MATLAB/Simulink to generate datasets. These datasets are generated using detailed models, including pulse-width modulation (PWM) in the converter/inverter, for all the components depicted in Fig. 1. It's important to note that generating data based on these detailed models through time-domain simulation can be a time-consuming process. As a result, the study opts to generate only orthogonal datasets, implemented through Taguchi's method [23], which helps streamline the dataset generation process.

Taguchi's method is a part of the broader field of design of experiments (DOE). DOE focuses on studying the effects of various control (design) factors on the performance of a system or process. When considering all possible combinations of these design factors, full factorial experiments would be conducted to design the experiment. However, Taguchi's method takes a more efficient approach by considering only orthogonal (minimal) experiments. By doing so, it reduces the number of experiments needed while still providing valuable insights. Taguchi's method can be characterized as a fractional factorial design, where it explores combinations of design factors at different given levels, further streamlining the experimental design process. Dr. Taguchi presented many orthogonal arrays denoted as L_x (y^z) where x, y and z represent the numbers of experiments, levels and design factors, respectively. For experiments in which all design factors have three levels, the array library consists of $L_9(3^4)$, $L_{27}(3^{13})$ and $L_{81}(3^{40})$ [23]. Table 2 illustrates the orthogonal array $L_9(3^4)$ encompassing four design factors (A, B, C and D), each of which has three levels (denoted as 1, 2 and 3); totally, only nine orthogonal experiments (L1∼L9) are conducted in case that the inputs are fixed [23]. However, if a problem has four design factors, each of which has three levels, then 3^4 (81) full factorial experiments are needed. Restated, out of 81 experiments, 72 experiments are mutually correlated to the nine orthogonal experiments in Table 2.

The design factors in this paper are irradiance (W/m^2), temperature (^{o}C) and short-circuit capacity (SCC in MVA) at the point of common coupling (PCC). Table 3 shows corresponding three levels of each design factors. Although there are only three design factors in the studied problem, the orthogonal array in Table 2 is still applicable by considering the 2nd, 3rd and 4th columns (design factors A, B and C only); these three columns still meet the property of orthogonality. Thus, nine orthogonal experiments are required to be performed by giving a fault condition and a mismatch.

TABLE 2. Orthogonal array $L_9(3^4)$.

experiments	A	B	C	D
L1	1	1	1	1
L2	1	2	2	2
L3	1	3	3	3
L4	2	1	2	3
L5	2	2	3	1
L6	2	3	1	2
L7	3	1	3	2
L8	3	2	1	3
L9	3	3	2	1

TABLE 3. Three levels of each design factor in training datasets.

Design factors	Level 1	Level 2	Level 3
A: irradiance (W/m^2)	300	600	1000
B: temperature (^{o}C)	10	25	40
C: SCC (MVA)	400	500	600

In Fig. 1, there are a total of 10 possible fault conditions: normal, line-line fault, open module in a string, shorted module in a string, open string, partial shading in a string, open modules in two strings, shorted modules in two strings, open strings, and partial shading in two strings. Beside, 20%, 40%, 60%, 80% and 100% mismatch conditions in a string are considered. In summary, the numbers of training datasets for the normal case and line-line fault case are 9×1 and 324, respectively; the number of training datasets for each other fault is 9×5 (that's, 9 orthogonal experiments multiplied with 5 mismatches). In total, 693 ($=9 \times 1 + 324 + 9 \times 5 \times 8$) training datasets are generated by MATLAB/Simulink.

In order to validate/test the well-trained proposed CCT model, other unseen testing datasets involving different levels of design factors are generated, as shown in Table 4. In total, 693 testing datasets are generated by MATLAB/Simulink.

TABLE 4. Three levels of each design factor in testing datasets.

Design factors	Level 1	Level 2	Level 3
A: irradiance (W/m^2)	400	750	850
B: temperature (^{o}C)	15	30	35
C: SCC (MVA)	450	525	575

B. CONVERSION OF 2D DATA TO HEATMAPS

The measured quantities include voltages and currents at the output of the DC/DC converter and the output of the

DC/AC inverters. The 2-dimensional data consists of eight columns: 2 DC voltages and 2 DC currents at the DC links of two PV strings, as well as 2 AC voltages and 2 AC currents at the output of the inverters of two PV strings. The 2-dimensional datasets comprise 256 × 3 rows, where 256 represents the sampled points in a cycle, and 3 is the number of cycles. Figure 3 illustrates the 2-dimensional input signals for the proposed method and other methods for comparison in Section V-D. This 2-dimensional data will be converted into an image, and the features will be further extracted by the proposed CCT model.

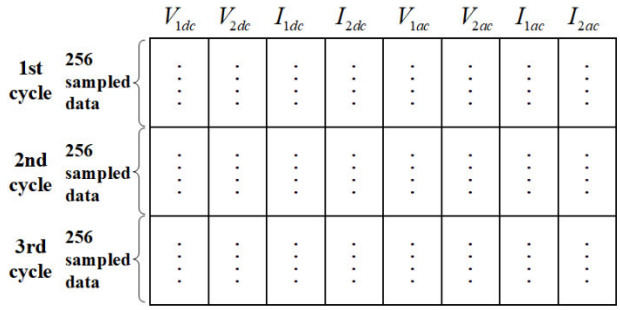


FIGURE 3. Two dimensional input data.

In the context of this study, the two-dimensional data obtained from the PV systems are transformed into images. Prior to this image conversion, the sampled data are standardized. This standardization involves adjusting the data using the corresponding mean values and standard deviations to ensure consistent and comparable data for further analysis. This process is essential to prepare the data for subsequent transformations and modeling, particularly when utilizing techniques inspired by transformer models with multi-head attention mechanisms, which have demonstrated success in computer vision studies.

In this study, heatmaps are employed with a size of 128 × 128 as vision images for the CCT model to classify PV faults. Heatmaps offer certain advantages compared to other techniques like Gramian Angular Field and recurrence plots, particularly in the context of vision applications using transformer models: (a) Interpretability: Heatmaps provide a visual representation of where the model is focusing its attention within an image. This interpretability can help in understanding which parts of the image are most important for the model's decision-making process. (b) Attention Visualization: Heatmaps allow for the visualization of the attention weights assigned by the transformer model to different regions of the input image. This visualization can provide insights into how the model processes and comprehends the image, shedding light on its decision-making process and highlighting key features or areas of interest.

A Python heatmap code from Seaborn library is used in this paper [25].

C. TUNING OF CCT MODEL BY PARTICLE SWARM OPTIMIZATION (PSO)

There are some parameters/hyperparameters in the proposed CCT model needed to be tuned optimally. Specifically, kernel size, kernel stride, the number of padding, pooling size, and pooling stride in the ConvLayer block. Besides, the number of transformer encoders and the number of multi-head attention are also required to be set optimally. To optimize these parameters and hyperparameters, this study employs particle swarm optimization (PSO). The optimization process aims to minimize the categorical cross-entropy, which is defined as the fitness function for the PSO. The reasons to adopt PSO are as follows: (a) The PSO in the Hyperactive library can handle integers, while most other optimization methods struggle with integer or discrete variables [26]. (b) Only three parameters (inertia factor, cognitive, and social factors; see appendix) need to be set, and more parameters may be required by other intelligent algorithms. (c) PSO does not require derivatives of the objective function, making it suitable for problems where the derivatives are not readily available or difficult to compute. The details of the PSO fundamental principles are provided in the appendix.

The search space for PSO is outlined in Table 5, detailing the range of values for each parameter/hyperparameters that the PSO algorithm will explore. To facilitate this optimization process, the study utilizes the PSO implementation available in the Python library Hyperactive.

TABLE 5. Search space of PSO.

Parameters/hyperparameters	Range
Kernel size	2, 3, 4
Kernel stride	1, 2, 3, 4
The number of padding	1, 2
Pooling size	2, 3, 4
Pooling stride	1, 2, 3, 4
The number of transformer encoder	2, 3, 4
The number of multi-head	1, 2, 3, 4

The work in [27] demonstrates that the results obtained by using PSO for integer programming—by rounding continuous variables to the nearest integers—are less sensitive to randomness characteristics compared to classical PSO, which deals with continuous variables. Restated, the standard deviation of the objective function obtained through multiple runs in an integer search space is reasonably small. In this paper, the best solution among five runs is adopted.

Please note that the output neurons of the proposed CCT model are set with 10 binary bits, each corresponding to one of the 10 classified conditions. These conditions include normality, partial shading, and 8 other types of PV faults. Only one of the 10 binary bits will be unity, and the others should be zero for each classification process. Compared to other existing methods, this paper addresses a broader range of PV fault types.

It is evident that all parameters/hyperparameters in Table 5 are directly associated with feature extraction, except for the number of padding, pooling size, and pooling stride. In comparison to traditional convolutional neural networks, the proposed CCT model integrates convolution operations using kernels and the transformer encoder, along with a multi-head attention mechanism, to effectively extract features from the data.

D. ALGORITHMIC STEPS OF PROPOSED METHOD

The proposed method can be implemented using the following algorithmic steps:

Step 1: Input data, including studied PV system, specific PV faults and mismatches.

Step 2: Define design factors and corresponding levels (see Table 3) of orthogonal experiments for producing training datasets.

Step 3: Generate training datasets based on normal conditions, PV faults and corresponding mismatches for each orthogonal experiment (see Table 2).

Step 4: Define unseen testing data (see Table 4) that are different from the training datasets obtained in Step 3.

Step 5: Define tuned parameters/hyperparameters of CCT model and their corresponding search space for PSO.

Step 6: Set the required parameters of PSO. Let iteration index itr be 1.

Step 7: Let particle size of PSO be Ps(12 herein), $p = 1, 2, \ldots, Ps$. Generate Ps particles. Let $p = 1$.

Step 8: Optimize the parameters (such as kernels for the convolution operation) of the CCT model for the pth particle using Adam optimizer [28]. $p = p+1$.

Step 9: If $p = Ps+1$, then itr=itr+1 and $p = 1$.

Step 10: If itr equals the specified maximum iteration number (1000 herein) of PSO, then output the optimal CCT model and go to Step 11; otherwise, go to Step 8.

Step 11: Evaluate the performance of the well-trained CCT model using unseen testing data obtained in Step 4.

It is evident that the proposed method consists of two optimization loops: the outer loop optimizes the kernel size, kernel stride, the number of padding, pooling size, pooling stride, the number of transformer encoders, and the number of multi-head attention using PSO, while the inner loop tunes the weights and biases in the CCT model using the Adam optimizer in Step 8. This paper employs categorical cross-entropy (CCE) as the objective function of the Adam optimizer because a classification problem is being studied. Each particle in PSO corresponds to a value of CCE by providing a set of the kernel size, kernel stride, the number of padding, pooling size, pooling stride, the number of transformer encoders, and the number of multi-head attention. Hence, the objective function of PSO is also CCE. For the case of single-label categorical classification, the CCE is defined as follows [29]:

$$\mathrm{CCE} = \frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M} y_m^k \log(h_\theta\,(u_m, k)) \qquad (3)$$

where u_m is the input for training example m. M and K are the numbers of training examples and classes, respectively. The symbol y_m^k is the target label for training example m for class k. h_θ represents the model with neural network weight θ.

A similar problem, specifically the fault diagnosis of an analog circuit, has been studied before [30], [31]. The analog circuits in question may include a bandpass filter or a nonlinear rectifier, typically constituting a single-input-single-output circuit. However, PV fault classification is more complex as it involves both DC and AC signal measurements, various levels of irradiance/temperature as inputs, and many PV strings in parallel. This work presents results using two PV strings, resulting in 8 outputs (4 DC and 4 AC measured signals). Therefore, an image-based deep learning model, such as CCT, with 2-dimensional images as inputs is necessary.

V. SIMULATION RESULTS

Research on diagnosis of PV array fault often involves the use of simulation data for several reasons: (a) Controlled environment: This level of control is often difficult to achieve in real-world scenarios where external factors (such as irradiance, temperature and fault location) can introduce additional complexity and variability. (b) Availability of data: Realistic fault data from PV systems may be limited or difficult to obtain. Simulations offer the advantage of reproducibility, allowing researchers to repeat/validate experiments under the same conditions. (c) Safety: Working with real PV systems can pose safety risks, especially when inducing faults or manipulating electrical components. Accordingly, a workstation with a Core(TM) i5-4460 CPU @ 3.20GHz, 64GB RAM and a NVIDIA GeForce RTX 3060 12G GPU is used to implement the proposed method. The training and testing datasets are generated using Matlab R2018b Simulink.

A. HEATMAPS GENERATED FROM VARIOUS CONDITIONS

In this paper, different heatmap image sizes (32×32, 64×64, and 128×128), various combinations of DC/AC signals (currents and voltages), and different numbers of signal cycles (1, 2, and 3) are investigated. Fig. 3 illustrates six heatmaps generated under different conditions: (a) normality, (b) line-line fault, (c) open module in a string, (d) shorted module in a string, (e) open string fault, and (f) partial shading, which involve 128×128 image size, both AC/DC currents and voltages and 3-cycle signals. As depicted in Fig. 4, the heatmaps exhibit significant differences among various fault conditions. These variations in the heatmaps are substantial and distinguishable, making them valuable for use in PV fault classifications. The unique patterns and features in these heatmaps can serve as a basis for effectively identifying and categorizing different types of faults in photovoltaic systems.

B. TUNING OF CCT MODEL BY PSO

Fig. 5 presents the variation of the fitness function (categorical cross-entropy) with respect to iterations during the tuning process of the CCT model using the PSO algorithm. The plot

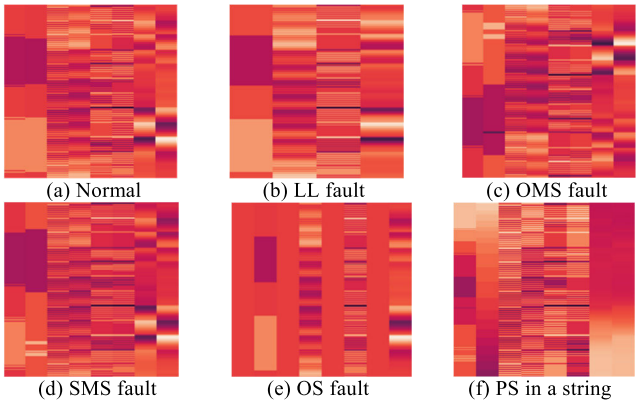


FIGURE 4. Heatmaps of 6 conditions.

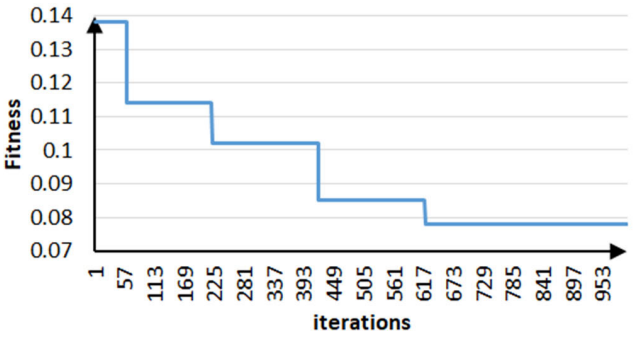


FIGURE 5. Fitness values with respect to iterations obtained by PSO.

demonstrates that the PSO algorithm requires approximately 620 iterations to converge to an optimal solution. This indicates that, over the course of these iterations, the algorithm fine-tunes the parameters and hyperparameters of the CCT model to minimize the categorical cross-entropy and achieve an optimal configuration for the PV fault classification task.

Fig. 6 illustrates the optimal CCT model obtained through the PSO process. This optimal model configuration is tailored for the use of 128 × 128 image size, considering both AC/DC currents and voltages, and utilizing 3-cycle signals. Specifically, the kernel size and pooling size are 3 × 3 and 4 × 4, respectively; the numbers of transformer encoder and multi-head are 4 and 2, respectively. The output of the model comprises 10 binary bits, each of which corresponds to one of the 10 conditions being classified. These conditions include: normality, partial shading and 8 other types of PV faults. Each binary bit represents the presence or absence of the respective condition, and the model's output is used to classify the input data into one of these 10 categories (see Sec. IV-A).

Seven possible CCT models that consider different image sizes, AC/DC currents and voltages and the numbers of cycle signals, as shown in Table 6, are investigated herein. The performance of these various models will be explored and compared in the next subsection. Table 7 shows these CCT models obtained by PSO.

TABLE 6. Seven CCT models involving various inputs.

	Image size	DC/AC	No. of cycles
M1	128	DC+AC	3
M2	128	DC	3
M3	128	AC	3
M4	32	DC+AC	3
M5	64	DC+AC	3
M6	128	DC+AC	1
M7	128	DC+AC	2

TABLE 7. Parameters/hyperparameters of seven CCT models.

	M1	M2	M3	M4	M5	M6	M7
Kernel size	3	2	4	2	3	2	2
Kernel stride	2	2	2	1	2	2	1
No. of padding	1	1	1	2	2	2	2
Pooling size	4	2	2	3	2	2	3
Pooling stride	3	3	2	3	3	2	3
No. of transformer encoder	4	1	1	4	1	3	4
No. of multi-head	2	3	4	3	4	3	3

C. ACCURACY OF DIFFERENT CCT MODLES

In this subsection, the performance of different CCT models is examined and summarized in Table 8.

TABLE 8. Performance of different CCT models.

	M1	M2	M3	M4	M5	M6	M7
Training accuracy	98.22%	93.14%	96.71%	94.43%	97.34%	94.63%	96.52%
Testing accuracy	96.82%	90.31%	95.23%	89.62%	94.74%	90.12%	94.51%
Training time	35:38:4	56:36:24	43: 54:12	28:06:08	31:31:43	44:50:25	52:48:34
Testing time	7.24s	7.37s	7.44s	7.31s	7.11s	7.21s	7.29s

Several key findings can be noted:

(i) Model M1, which considers an image size of 128 × 128, both AC/DC currents and voltages, and 3-cycle signals, achieves the best training accuracy at 98.22% and testing accuracy at 97.34%.

(ii) Model M4, which uses an image size of 32 × 32, both AC/DC currents and voltages, and 3-cycle signals, results in the lowest testing accuracy at 89.62%. This suggests that a smaller image size reduces resolution, potentially causing some signal features to be overlooked, leading to lower accuracy.

(iii) The process of tuning the CCT model using the PSO simultaneously involves the adjustment of various parameters, including the values of kernels, which are optimized by the Adam optimizer. Model M4 is the fastest in this optimization process, while model M2 takes the longest time. This indicates that the complexity of the model and the size of the input data influence the tuning time required.

(iv) Despite the relatively long training time, the testing time for a single case (a single image) is quite efficient, requiring only about 7 seconds. It's worth noting that the

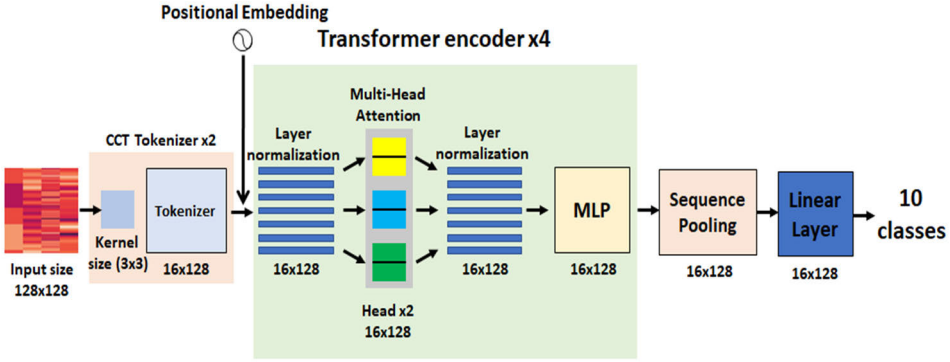


FIGURE 6. Optimal CCT model (128 × 128 image size, both AC/DC currents and voltages and 3-cycle signals).

Python heatmap code from the Seaborn library [25] takes approximately 0.03 seconds to convert a single dataset into a heatmap. This efficiency in testing time is valuable for real-time or near-real-time applications, as it ensures quick response when classifying PV faults based on incoming data.

D. COMPARATIVE STUDIES

This research conducts two comparative studies to evaluate the proposed method against: (a) CNN-based methods, including AlexNet [32], ResNet50 [33], VGG16 [34], VGG19 [34], and GoogleNet [35] and (b) traditional machine learning algorithms, which consist of Support Vector Machine (SVM), Decision Tree (DT), k-nearest neighbor (kNN), and Random Forest (RF). Especially, SVM, DT, and kNN have become benchmark methods frequently used in other comparative studies, such as the classification of power transformer faults [36].

For these comparative studies, all methods utilize standardized heatmaps associated with an image size of 128 × 128 as input data, while the outputs of all methods consist of 10 binary bits signifying 10 classes. This standardized input incorporates both AC and DC currents and voltages and is based on 3-cycle signals. The parameters of the AlexNet, ResNet50, VGG16, VGG19, and GoogleNet models are consistent with those specified in [32], [33], [34], and [35] because those settings were well-tuned to achieve the best results.

Table 9 presents the comparative results between the proposed method and various CNN-based methods, and several key findings can be summarized:

(i) The proposed method achieves the highest training accuracy (98.22%) and testing accuracy (97.34%) when compared to the other CNN-based methods.

(ii) AlexNet exhibits the lowest training accuracy (80.75%) and testing accuracy (76.47%) among all the methods.

(iii) To ensure a fair comparison, the training time for the proposed method, as indicated in Table 9, is obtained based on the results from the PSO. Under this condition, the proposed method requires the shortest training time (1:24:02), while ResNet50 demands the longest training time (4:36:02).

TABLE 9. Comparative results between proposed method and CNN-based methods.

Methods	Training accuracy	Testing accuracy	Training time	Testing time
Proposed Method	98.22%	96.82%	1:24:02	7.24s
AlexNet	80.75%	76.47%	1:54:42	0.53s
GoogleNet	83.27%	80.14%	3:48:11	0.77s
ResNet50	95.31%	93.23%	4:36:02	1.12s
VGG16	92.76%	90.11%	2:30:37	0.81s
VGG19	90.14%	87.35%	3:01:21	0.62s

(iv) With respect to testing time for a given dataset, the proposed method has the longest average testing time (7.24 s), while AlexNet is notably faster with a shorter testing time (0.53 s).

(v) When considering the two most crucial factors, testing accuracy and training time, the proposed method outperforms the second-best method, ResNet50.

These findings underscore the effectiveness and efficiency of the proposed method in terms of both testing accuracy and training time compared to other CNN-based methods.

Table 10 presents the comparative results between the proposed method and classical machine learning methods, such as SVM, DT, kNN and RF. The parameters of the SVM, DT, kNN, and RF models are configured based on the guidelines provided in [9] to ensure a fair comparison. Several key findings can be summarized:

TABLE 10. Comparative results between proposed method and classical machine learning methods.

Methods	Training accuracy	Testing accuracy	Training time	Testing time
Proposed Method	98.22%	96.82%	1:24:02	7.24s
SVM	92.28%	80.62%	1.29s	1.17s
DT	91.36%	81.65%	5.86s	1.85s
kNN	88.07%	80.11%	1.60s	1.23s
RF	90.48%	84.24%	3.82s	1.02s

(i) The proposed method achieves the highest training accuracy (98.22%) and testing accuracy (97.34%) when compared to other classical machine learning methods.

(ii) kNN attains the lowest training accuracy (88.07%) and testing accuracy (80.11%) among all the methods.

(iii) Classical machine learning methods generally require significantly shorter training times compared to the deep learning-based proposed method.

These findings demonstrate the superior performance of the proposed method in terms of both training and testing accuracy when compared to classical machine learning methods, despite the longer training times associated with deep learning techniques.

VI. CONCLUSION

This paper introduces a novel method based on the Compact Convolutional Transformer (CCT) for classifying different types of photovoltaic faults. The essential contributions and findings are summarized as follows:

(i) Orthogonal Datasets: The proposed method uses orthogonal datasets for training the CCT model, addressing the challenge of data scarcity common in classical convolution-based transformer models. This approach significantly reduces computational demands while maintaining competitive testing accuracy.

(ii) Parameter Optimization: The method employs two iterative loops to optimize the parameters, hyperparameters, and weights of the CCT model. This eliminates the need for heuristic tuning, resulting in favorable results.

(iii) Dataset Features: The datasets incorporate a 128×128 image size, include both AC and DC currents and voltages, and span three cycles of signals. This rich dataset feature set contributes to achieving high accuracy in fault classification.

(iv) Performance Comparison: The proposed method outperforms both CNN-based methods and classical machine learning methods in terms of testing accuracy.

(v) Real-Time Application: The testing time for an image is approximately 7.24 seconds, making the method suitable for real-time applications.

In future work, the authors plan to expand their research to encompass a comprehensive four-step diagnosis process: detection, localization, classification, and remedy. They intend to leverage state-of-the-art techniques such as digital twin modeling and edge computing to enhance the diagnosis processes. This indicates a commitment to further advancing the field of fault diagnosis in photovoltaic systems.

APPENDIX

PSO encompasses two updating formulas for the position of the ith particle.

$$V_i(k+1) = w \times V_i(k) + c_1\theta_1 \left(P_{\text{ibest}}(k) - P_i(k)\right) + c_2\theta_2 \left(Gbest(k) - P_i(k)\right). \quad \text{(A-1)}$$

$$P_i(k+1) = P_i(k) + V_i(k+1), \quad \text{(A-2)}$$

where w is the inertia factor; c_1 and c_2 are the cognitive and social factors, and θ_1 and θ_2 are random numbers in a range of [0, 1]. Equations (A-1) and (A-2) are the velocity and position of the ith particle at time $(k+1)$. The best overall solution of all the particles in the population is known as Gbest (k) and $P_{\text{ibest}}(k)$ is the best known position of the ith particle.

$P_i(k)$ is a vector consisting of components (all integers) as shown in Table 5 in this paper. Each particle of the swarm is then rounded to the closest integer after determining its new position using (A-2).

REFERENCES

[1] *Climate Change: Global Temperature*. Accessed: Aug. 10, 2023. [Online]. Available: https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

[2] F. M. Aboshady and I. B. M. Taha, "Fault detection and classification scheme for PV system using array power and cross-strings differential currents," *IEEE Access*, vol. 9, pp. 112655–112669, 2021, doi: 10.1109/ACCESS.2021.3104007.

[3] B. K. Karmakar and A. K. Pradhan, "Detection and classification of faults in solar PV array using Thevenin equivalent resistance," *IEEE J. Photovolt.*, vol. 10, no. 2, pp. 644–654, Mar. 2020, doi: 10.1109/JPHOTOV.2019.2959951.

[4] C. Li, Y. Yang, F. Fan, L. Xia, P. Peng, Y. Wang, K. Zhang, and H. Wei, "A novel methodology for partial shading diagnosis using the electrical parameters of photovoltaic strings," *IEEE J. Photovolt.*, vol. 12, no. 4, pp. 1027–1035, Jul. 2022, doi: 10.1109/JPHOTOV.2022.3173723.

[5] L. Xu, Z. Pan, C. Liang, and M. Lu, "A fault diagnosis method for PV arrays based on new feature extraction and improved the fuzzy C-mean clustering," *IEEE J. Photovolt.*, vol. 12, no. 3, pp. 833–843, May 2022, doi: 10.1109/JPHOTOV.2022.3151330.

[6] A. Ul-Haq, H. F. Sindi, S. Gul, and M. Jalal, "Modeling and fault categorization in thin-film and crystalline PV arrays through multilayer neural network algorithm," *IEEE Access*, vol. 8, pp. 102235–102255, 2020, doi: 10.1109/ACCESS.2020.2996969.

[7] H. Zhu, S. A. Z. Ahmed, M. A. Alfakih, M. A. Abdelbaky, A. R. Sayed, and M. A. A. Saif, "Photovoltaic failure diagnosis using sequential probabilistic neural network model," *IEEE Access*, vol. 8, pp. 220507–220522, 2020, doi: 10.1109/ACCESS.2020.3043129.

[8] M. Laurino, M. Piliougine, and G. Spagnuolo, "Artificial neural network based photovoltaic module diagnosis by current–voltage curve classification," *Sol. Energy*, vol. 236, pp. 383–392, Apr. 2022, doi: 10.1016/j.solener.2022.02.039.

[9] B. Li, C. Delpha, A. Migan-Dubois, and D. Diallo, "Fault diagnosis of photovoltaic panels using full I–V characteristics and machine learning techniques," *Energy Convers. Manage.*, vol. 248, Nov. 2021, Art. no. 114785, doi: 10.1016/j.enconman.2021.114785.

[10] M. M. Badr, M. S. Hamad, A. S. Abdel-Khalik, R. A. Hamdy, S. Ahmed, and E. Hamdan, "Fault identification of photovoltaic array based on machine learning classifiers," *IEEE Access*, vol. 9, pp. 159113–159132, 2021, doi: 10.1109/ACCESS.2021.3130889.

[11] A. Eskandari, J. Milimonfared, and M. Aghaei, "Fault detection and classification for photovoltaic systems based on hierarchical classification and machine learning technique," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12750–12759, Dec. 2021, doi: 10.1109/TIE.2020.3047066.

[12] S. Rao, G. Muniraju, C. Tepedelenlioglu, D. Srinivasan, G. Tamizhmani, and A. Spanias, "Dropout and pruned neural networks for fault classification in photovoltaic arrays," *IEEE Access*, vol. 9, pp. 120034–120042, 2021, doi: 10.1109/ACCESS.2021.3108684.

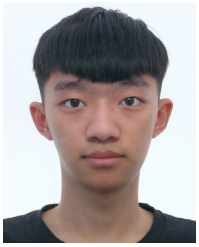
[13] S.-Q. Chen, G.-J. Yang, W. Gao, and M.-F. Guo, "Photovoltaic fault diagnosis via semisupervised ladder network with string voltage and current measures," *IEEE J. Photovolt.*, vol. 11, no. 1, pp. 219–231, Jan. 2021, doi: 10.1109/JPHOTOV.2020.3038335.

[14] K. Dhibi, M. Mansouri, K. Bouzrara, H. Nounou, and M. Nounou, "An enhanced ensemble learning-based fault detection and diagnosis for grid-connected PV systems," *IEEE Access*, vol. 9, pp. 155622–155633, 2021, doi: 10.1109/ACCESS.2021.3128749.

[15] R. H. F. Alves, G. A. D. Deus, Jr., E. G. Marra, and R. P. Lemos, "Automatic fault classification in photovoltaic modules using convolutional neural networks," *Renew. Energy*, vol. 179, pp. 502–516, Dec. 2021, doi: 10.1016/j.renene.2021.07.070.

[16] F. Aziz, A. Ul Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*, vol. 8, pp. 41889–41904, 2020, doi: 10.1109/ACCESS.2020.2977116.
[17] W. Gao and R.-J. Wai, "A novel fault identification method for photovoltaic array via convolutional neural network and residual gated recurrent unit," *IEEE Access*, vol. 8, pp. 159493–159510, 2020, doi: 10.1109/ACCESS.2020.3020296.
[18] M. Alrifaey, W. H. Lim, C. K. Ang, E. Natarajan, M. I. Solihin, M. R. M. Juhari, and S. S. Tiang, "Hybrid deep learning model for fault detection and classification of grid-connected photovoltaic system," *IEEE Access*, vol. 10, pp. 13852–13869, 2022, doi: 10.1109/ACCESS.2022.3140287.
[19] A. Eskandari, M. Aghaei, J. Milimonfared, and A. Nedaei, "A weighted ensemble learning-based autonomous fault diagnosis method for photovoltaic systems using genetic algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 144, Jan. 2023, Art. no. 108591, doi: 10.1016/j.ijepes.2022.108591.
[20] H. Guo, S. Hu, F. Wang, and L. Zhang, "A novel method for quantitative fault diagnosis of photovoltaic systems based on data-driven," *Electr. Power Syst. Res.*, vol. 210, Sep. 2022, Art. no. 108121, doi: 10.1016/j.epsr.2022.108121.
[21] A. Vaswani, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 6000–6010.
[22] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," 2021, *arXiv:2104.05704*.
[23] G. Taguchi, S. Chowdhury, and S. Taguchi, *Robust Engineering*. London, U.K.: McGraw-Hill, 2000.
[24] N.M. Vacanti, "The fundamentals constructing interpreting heat maps," in *Metabolic Signaling. Methods in Molecular Biology*, vol. 1862, S. M. Fendt and S. Lunt, Eds. New York, NY, USA: Humana Press, 2019, doi: 10.1007/978-1-4939-8769-6_20.
[25] *seaborn.heatmap*. Accessed: Dec. 10, 2022. [Online]. Available: seaborn.pydata.org/generated/seaborn.heatmap.html
[26] *hyperactive 4.6.0*. Accessed: Oct. 24, 2023. [Online]. Available: https://pypi.org/project/hyperactive/
[27] E. C. Laskari, K. E. Parsopoulos, and M. N. Vrahatis, "Particle swarm optimization for integer programming," in *Proc. Congr. Evol. Computation (CEC)*, Honolulu, HI, USA, May 2002, pp. 1582–1587.
[28] A. Gulli and S. Pal, *Deep Learning With Keras*. Birmingham, U.K.: Packt Publishing, 2017.
[29] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
[30] C. Zhang, Y. He, T. Yang, B. Zhang, and J. Wu, "An analog circuit fault diagnosis approach based on improved wavelet transform and MKELM," *Circuits, Syst., Signal Process.*, vol. 41, no. 3, pp. 1255–1286, Mar. 2022.
[31] C. Zhang, Y. He, L. Zuo, J. Wang, and W. He, "A novel approach to diagnosis of analog circuit incipient faults based on KECA and OAO LSSVM," *Metrology Meas. Syst.*, vol. 22, no. 2, pp. 251–262, Jun. 2015.
[32] A. Rasheed, N. Ali, B. Zafar, A. Shabbir, M. Sajid, and M. T. Mahmood, "Handwritten Urdu characters and digits recognition using transfer learning and augmentation with AlexNet," *IEEE Access*, vol. 10, pp. 102629–102645, 2022, doi: 10.1109/ACCESS.2022.3208959.
[33] H. Asma-Ull, I. D. Yun, and B. L. Yun, "Regression to classification: Ordinal prediction of calcified vessels using customized ResNet50," *IEEE Access*, vol. 11, pp. 48783–48796, 2023, doi: 10.1109/ACCESS.2023.3270562.
[34] S. A. P. N. Kavala and R. Pothuraju, "Detection of grape leaf disease using transfer learning methods: VGG16 & VGG19," in *Proc. 6th Int. Conf. Comput. Methodolog. Commun. (ICCMC)*, Erode, India, Mar. 2022, pp. 1205–1208, doi: 10.1109/ICCMC53470.2022.9753773.
[35] A. Sekhar, S. Biswas, R. Hazra, A. K. Sunaniya, A. Mukherjee, and L. Yang, "Brain tumor classification using fine-tuned GoogLeNet features and machine learning algorithms: IoMT enabled CAD system," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 983–991, Mar. 2022, doi: 10.1109/JBHI.2021.3100758.
[36] M. Bigdeli, P. Siano, and H. H. Alhelou, "Intelligent classifiers in distinguishing transformer faults using frequency response analysis," *IEEE Access*, vol. 9, pp. 13981–13991, 2021.

YING-YI HONG (Senior Member, IEEE) received the B.S.E.E. degree from Chung Yuan Christian University (CYCU), Taiwan, in 1984, the M.S.E.E. degree from National Cheng Kung University (NCKU), Taiwan, in 1986, and the Ph.D. degree from the Department of Electrical Engineering, National Tsing Hua University (NTHU), Taiwan, in December 1990. Sponsored by the Ministry of Education, Taiwan, he conducted research with the Department of Electrical Engineering, University of Washington, Seattle, WA, USA, from August 1989 to August 1990. He has been with CYCU, since 1991. He was the Dean of the College of Electrical Engineering and Computer Science, CYCU, from 2006 to 2012. He was promoted to be a Distinguished Professor due to his exceptional research, leadership, teamwork, and international collaboration, in 2012. From 2012 to 2018, he was a Secretary General with CYCU. He is currently the Vice President of CYCU. His research interests include power system analysis and artificial intelligence applications. He received the Outstanding Professor of Electrical Engineering Award from the Chinese Institute of Electrical Engineering (CIEE), Taiwan, in 2006. He was the Chair of the IEEE PES Taipei Chapter, in 2001.



LI-FAN CHEN received the B.S.E.E. degree from the Chien Hsin University of Science and Technology, in 2021, and the M.S.E.E. degree from Chung Yuan Christian University, in 2023. His research interests include the applications of artificial intelligence, microgrid analysis, evolutionary algorithms, and quantum computing.

WEINA ZHANG received the M.S. degree in system analysis and integration and the Ph.D. degree in computer science and technology from the South China University of Technology (SCUT), Guangzhou, China, in 2008 and 2020, respectively. She is currently a Lecturer with the College of Computer Science and Technology, Shanghai University of Electric Power. Her research interests include recommendation algorithms with interpretability, stability, and security, power load forecasting, and fault diagnosis.